



## King's Research Portal

DOI:

[10.1016/j.neurobiolaging.2016.04.004](https://doi.org/10.1016/j.neurobiolaging.2016.04.004)

*Document Version*

Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Sassi, C., Nalls, M. A., Ridge, P. G., Gibbs, R., Ding, J., Lupton, M. K., Troakes, C., Lunnon, K., Al-Sarraj, S., Brown, K. S., Medway, C., Clement, N., Lord, J., Turton, J., Bras, J., Almeida, M. R., Passmore, P., Craig, D., Johnston, J., ... Hardy, J. (2016). ABCA7 p.G215S as potential protective factor for Alzheimer's disease. *Neurobiology of Aging*. <https://doi.org/10.1016/j.neurobiolaging.2016.04.004>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

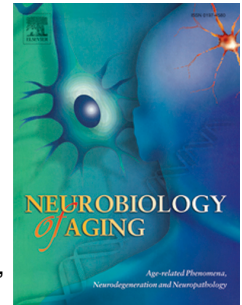
### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Accepted Manuscript

ABCA7 p.G215S as potential protective factor for Alzheimer's disease

C. Sassi, M.A. Nalls, P.G. Ridge, R. Gibbs, J. Ding, M.K. Lupton, C. Troakes, K. Lunnon, S. Al-Sarraj, K.S. Brown, C. Medway, N. Clement, J. Lord, J. Turton, J. Bras, M.R. Almeida, ARUK Consortium, H. Holstege, E. Louwersheimer, W. van der Flier, P. Scheltens, J.C. Van Swieten, I. Santana, C. Oliveira, K. Morgan, J.F. Powell, J.S. Kauwe, C. Cruchaga, A.M. Goate, A.B. Singleton, R. Guerreiro, J. Hardy



PII: S0197-4580(16)30031-8

DOI: [10.1016/j.neurobiolaging.2016.04.004](https://doi.org/10.1016/j.neurobiolaging.2016.04.004)

Reference: NBA 9582

To appear in: *Neurobiology of Aging*

Received Date: 9 January 2016

Revised Date: 23 March 2016

Accepted Date: 10 April 2016

Please cite this article as: Sassi, C., Nalls, M.A., Ridge, P.G., Gibbs, R., Ding, J., Lupton, M.K., Troakes, C., Lunnon, K., Al-Sarraj, S., Brown, K.S., Medway, C., Clement, N., Lord, J., Turton, J., Bras, J., Almeida, M., ARUK Consortium, Holstege, H., Louwersheimer, E., van der Flier, W., Scheltens, P., Van Swieten, J.C., Santana, I., Oliveira, C., Morgan, K., Powell, J.F., Kauwe, J.S., Cruchaga, C., Goate, A., Singleton, A.B., Guerreiro, R., Hardy, J., ABCA7 p.G215S as potential protective factor for Alzheimer's disease, *Neurobiology of Aging* (2016), doi: 10.1016/j.neurobiolaging.2016.04.004.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# *ABCA7 p.G215S as potential protective factor for Alzheimer's disease*

Sassi C.<sup>1, 2, 3, 4</sup>, Nalls M.A.<sup>2</sup>, Ridge P.G.<sup>5</sup>, Gibbs R.<sup>2</sup>, Ding J.<sup>2</sup>, Lupton M.K.<sup>6</sup>, Troakes C.<sup>6</sup>, Lunnon K.<sup>6</sup>, Al-Sarraj S.<sup>6</sup>, Brown K.S.<sup>7</sup>, Medway C.<sup>7</sup>, Clement N.<sup>7</sup>, Lord J.<sup>7</sup>, Turton J.<sup>7</sup>, Bras J.<sup>1</sup>, Almeida MR.<sup>8</sup>, ARUK Consortium\*, Holstege H.<sup>9</sup>, Louwersheimer E.<sup>9</sup>, van der Flier W.<sup>9</sup>, Scheltens P.<sup>9</sup>, Van Swieten J.C.<sup>10</sup>, Santana I.<sup>11</sup>, Oliveira C.<sup>12</sup>, Morgan K.<sup>5</sup>, Powell J.F.<sup>4</sup>, Kauwe J.S.<sup>3</sup>, Cruchaga C.<sup>13</sup>, Goate AM.<sup>14</sup>, Singleton A.B.<sup>2</sup>, Guerreiro R.<sup>1</sup>, Hardy J.<sup>1</sup>

\*The Alzheimer's Research UK (ARUK) Consortium: Peter Passmore, David Craig, Janet Johnston, Bernadette McGuinness, Stephen Todd, Queen's University Belfast, UK; Reinhard Heun, Royal Derby Hospital, UK; Heike Kölsch, University of Bonn, Germany; Patrick G. Kehoe, University of Bristol, UK; Emma R.L.C. Vardy, Salford Royal NHS Foundation Trust, UK; Nigel M. Hooper, David M. Mann, Stuart Pickering-Brown, University of Manchester, UK; Kristelle Brown, James Lowe, Kevin Morgan, University of Nottingham, UK; A. David Smith, Gordon Wilcock, Donald Warden, University of Oxford (OPTIMA), UK; Clive Holmes, University of Southampton, UK.

<sup>1</sup> Reta Lila, Weston Research Laboratories, Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK;

<sup>2</sup> Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA;

<sup>3</sup> Department of Experimental Neurology, Center for Stroke Research Berlin (CSB), Charité' Universitätsmedizin, Berlin, Germany

<sup>4</sup> German Center for Neurodegenerative Diseases (DZNE), Berlin site, Germany

<sup>5</sup> Departments of Biology, Neuroscience, Brigham Young University, Provo, UT, USA

<sup>6</sup> King's College London Institute of Psychiatry, London, UK

<sup>7</sup> Translation Cell Sciences-Human Genetics, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK

<sup>8</sup> Neurogenetics Laboratory, Center for Neuroscience and Cell Biology, University of Coimbra, 3004-548 Coimbra, Portugal

<sup>9</sup> Alzheimer Center, Department of Neurology, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam 1081 HZ, The Netherlands

<sup>10</sup> Alzheimer Center, Department of Neurology, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam 1081 HZ, The Netherlands

<sup>11</sup> Neurology Department, Centro Hospitalar e Universitário de Coimbra, 3004-548 Coimbra, Portugal and Faculty of Medicine, Coimbra University, 3004-548 Coimbra, Portugal

<sup>12</sup> CNC - Center for Neuroscience and Cell Biology, University of Coimbra, Portugal; Laboratory of Biochemistry, Faculty of Medicine, University of Coimbra, Portugal

<sup>13</sup> Washington University, Division of Biology and Biomedical Sciences St. Louis, MO, USA

<sup>14</sup> Icahn School of Medicine at Mount Sinai, Icahn Medical Institute, New York, NY, USA

**Corresponding authors****Prof. John Hardy**

Reta Lila, Weston Research Laboratories, Department of Molecular Neuroscience, UCL  
Institute of Neurology, London, UK

j.hardy@ucl.ac.uk

**Dr. Celeste Sassi**

Reta Lila, Weston Research Laboratories, Department of Molecular Neuroscience, UCL  
Institute of Neurology, London, UK

celeste.sassi.10@alumni.ucl.ac.uk

**ABSTRACT**

Genome-wide association studies (GWASs) have been effective approaches to dissect common genetic variability underlying complex diseases in a systematic and unbiased way. Recently, GWASs have led to the discovery of over 20 susceptibility loci for Alzheimer's disease (AD).

Despite the evidence showing the contribution of these loci to AD pathogenesis, their genetic architecture has not been extensively investigated, leaving the possibility that low frequency and rare coding variants may also occur and contribute to the risk of disease. We have used exome and genome sequencing data to analyse the single independent and joint effect of rare and low frequency protein coding variants in 9 AD GWAS loci with the strongest effect sizes after *APOE* (*BIN1*, *CLU*, *CR1*, *PICALM*, *MS4A6A*, *ABCA7*, *EPHA1*, *CD33*, *CD2AP*) in a cohort of 332 sporadic AD cases and 676 elderly controls of British and North American ancestry. We identified coding variability in *ABCA7* as contributing to AD risk. This locus harbors a low frequency coding variant (p.G215S, rs72973581, MAF=4.3%) conferring a modest but statistically significant protection against AD (p-value=  $6 \times 10^{-4}$ , OR= 0.57, 95% CI 0.41-0.80). Notably, our results are not driven by an enrichment of loss of function variants in *ABCA7*, recently reported as main pathogenic factor underlying AD risk at this locus. In summary, our study confirms the role of *ABCA7* in AD and provide new insights that should address functional studies.

**KEYWORDS**

Alzheimer's disease (AD); Genome-wide association studies (GWASs); *ABCA7*; whole exome sequencing (WES); whole genome sequencing (WGS); protective variant

**1. INTRODUCTION**

Alzheimer's disease (AD) is the most common cause of progressive dementia in the elderly. Aging and genetic factors play a critical role for the disease development. Rare coding and fully penetrant mutations in *APP*, *PSEN1* and *PSEN2* explain part of the AD autosomal dominant cases. On the other hand, *APOE*  $\epsilon 4$  allele and rare coding variants in *TREM2*

represent the main risk factors for late-onset and apparently sporadic AD (Chartier-Harlin et al., 1994) (Guerreiro et al., 2013). In the last 5 years, genome-wide association studies (GWASs) identified over 20 main risk loci influencing AD susceptibility (Harold et al., 2009)(Hollingworth et al., 2011)(Lambert et al., 2009)(Naj et al., 2011)(Seshadri et al., 2010)(Lambert et al., 2013). Among these, 9 have been replicated by at least two independent GWASs and present the strongest effect sizes after *APOE* (*BIN1*, *CLU*, *CR1*, *PICALM*, *MS4A6A*, *ABCA7*, *EPHA1*, *CD33*, *CD2AP*).

GWASs have been a successful strategy to identify loci associated to a common trait, shedding light on disease pathways, and for AD these include 1) immune response (*CR1*, *MSA4A/MSA7A*, *CD2AP*, *CD33*, *EPHA1*, *ABCA7*), 2) vesicles trafficking (*PICALM*, *BIN1*), 3) lipid metabolism (*CLU*, *ABCA7*) and 4) A $\beta$  peripheral clearance (*PICALM*, *BIN1*, *CD33* and *ABCA7*). (<http://www.alzgene.org/>). Nevertheless, the functional variant(s) within these risk loci have not yet been fully defined.

GWAS arrays tag common, low penetrant and generally non coding variants that likely exert a subtle regulatory effect ( $0.8 < OR < 1.5$ ) on a trait, affecting gene expression, CpG islands methylation and splicing, in *cis* or *trans* (Visscher et al., 2012)(Ramasamy et al., 2014). Whereas, low frequency ( $1\% < MAF < 5\%$ ) and rare variants ( $MAF < 1\%$ ) with a modest penetrance remain mostly undetected either because they are not in the array or because, even with the implementation of imputation, the detection of variants with  $MAF < 2\%$  is not sufficiently accurate. As an illustrative example, *APOE* GWAS hit maps to an intronic region and it is likely driven by the *APOE*  $\epsilon 4$  allele, which is a common coding haplotype (rs429358, p.C130R and rs7412, p.R176R  $MAF=15\%$ ) in exon 4, that it is not tagged by the custom genotyping arrays mostly used.

Recently, resequencing studies have been powerful strategies to bridge the gap between susceptibility loci identified and actual disease-modifying variant(s) (Beaudoin et al., 2013)(Rivas et al., 2011)(Service et al., 2014)(Lohmueller et al., 2013).

Therefore, we have used exome and genome sequencing data 1) to identify rare and low frequency coding variants in *BIN1*, *CLU*, *CR1*, *PICALM*, *MS4A6A*, *ABCA7*, *EPHA1*, *CD33*, *CD2AP* and 2) to investigate their single independent and combined effect on AD susceptibility. Both the single-variant and the gene-based association tests confirmed *ABCA7* as

susceptibility locus associated with AD. Importantly, although *ABCA7* loss of function (LoF) mutations (indels, nonsense, splice-site mutations) have been recently reported as main mechanism increasing AD risk at this locus (Steinberg et al., 2015), our results are not driven by such variants. Whereas, we report an enrichment for *ABCA7* common and low frequency coding variants with a potential protective effect, that is mainly responsible for our gene-based signal. Among these, *ABCA7* p.G215S is the main low frequency missense hit in the single-variant analysis in the discovery cohort. The potential protective role of this variant has been further confirmed in an independent European and North American cohort. Our results show that *ABCA7* p.G215S exerts a mild but statistically significant influence, lowering the risk for AD. Thus, confirming *ABCA7* to be a good potential target to address functional studies.

## 2. MATERIALS AND METHODS

The discovery cohort was composed of 332 apparently sporadic AD cases and 676 elderly controls, neuropathologically and clinically confirmed, originating from the UK and North America. The mean age at disease onset was 71.66 years (range 41-94 years) for cases and the mean age of ascertainment was 78.15 years (range 60-102 years) for controls (**Table 1**). The majority of the cases (77%) were late-onset (> 65 years at onset).

Among the cases and controls, 42% and 51% were female, respectively. 58% and 47% of the cases and controls carried the *APOE*  $\epsilon$ 4 allele, respectively. The *APOE*  $\epsilon$ 4 allele was significantly associated to the disease status in the NIH and ADNI series (p-value= 0.02 and  $1.19 \times 10^{-9}$ , respectively). The threshold call rate for inclusion of the subject in analysis was 95%. On this cohort we performed 1) gene-based analysis (SKAT and c-alpha tests) and 2) single-variant association analysis, targeting 23.5 Kilobase pairs (Kbs) of coding sequence. Finally, we followed-up, in an independent Caucasian dataset, *ABCA7* p. G215S, the only nominal significant low frequency missense variant in the single-marker association test in our discovery set (**Figure 1**).

The follow-up dataset was composed of 307 late-onset apparently sporadic AD cases from North America and Europe and 501 elderly Caucasian controls from North America (Coriell repositories), Europe, Australia and Canada (**Table 1**). Written informed consent was

obtained for each clinically assessed individual and the study was approved by the appropriate institutional review boards. All samples had fully informed consent for retrieval and were authorized for ethically approved scientific investigation (UCLH Research Ethics Committee number 10/H0716/3, BYU IRB, Cardiff REC for Wales 08/MRE09/38+5, REC Reference 04/Q2404/130, National Research Ethics Service [NRES]).

### **2.1 Exome sequencing**

DNA was extracted from blood or brain for cases and brain only for controls using standard protocols. Library preparation for next generation sequencing used DNA (between 1 µg and 3 µg) fragmented in a Covaris E210 (Covaris Inc.). DNA was end-repaired by 5'phosphorylation, using the Klenow polymerase. A poly-adenine tail was added to the 3'end of the phosphorylated fragment and ligated to Illumina adapters. After purification using an AMPure DNA Purification kit (Beckman Coulter, Inc), adapter-ligated products were amplified. The DNA library was then hybridized to an exome capture library (NimbleGen SeqCap EZ Exome v2.0, Roche Nimblegen Inc. or TruSeq, Illumina Inc.) and precipitated using streptavidin-coated magnetic beads (Dyna Magnetic Beads, Invitrogen). These exome libraries were PCR-amplified, and then DNA hybridized to paired-end flow cells using a cBot (Illumina, Inc.) cluster generation system. Samples were sequenced on the Illumina HiSeq™ 2000 using 2x100 paired end reads cycles.

### **2.2 Whole Genome sequencing**

Genome sequencing was performed in 199 controls, from the Cache County Study on Memory in Aging. All samples were sequenced with the use of Illumina HiSeq technology.

### **2.3 Sanger sequencing**

*ABCA7* p. G215S (rs72973581) was screened in an additional follow-up cohort composed of 307 late-onset cases and 501 elderly controls. Primers were designed in Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) using UCSC (<http://genome.ucsc.edu/>) reference sequences NM\_019112 (*ABCA7*).



Purified sequences were analysed on an ABI 3730 DNA Analyzer (Applied Biosystems, CA, USA) and chromatograms were visualized in Sequencher software (version 4.2 Gene Codes Corporation, MI, USA).

## 2.4 Bioinformatics

Sequence alignment and variant calling were performed against the reference human genome (UCSC hg19). Alignment was performed with the use of CASAVA software and variant calling was performed with the use of SAMtools (Li et al., 2009) and GATK (McKenna et al., 2010). Paired end sequence reads (2x100bp paired end read cycles) were aligned using the Burrows-Wheeler aligner (Li and Durbin, 2009). Format conversion and indexing were performed with Picard ([www.picard.sourceforge.net/index.shtml](http://www.picard.sourceforge.net/index.shtml)). GATK was used to recalibrate base quality scores, perform local re-alignments around indels and to call and filter the variants (McKenna et al., 2010). VCFtools was used to annotate gene information for the remaining novel variants. We used ANNOVAR software to annotate the variants (Wang et al., 2010). Variants were checked against established databases (1000 Genomes Project and dbSNP v.134). Calling algorithms, pipe-lines and reference panels were the same the pooled datasets. The protein coding effects of variants was predicted using SIFT, Polyphen2 and SeattleSeq Annotation ([gvs.gs.washington.edu/SeattleSeqAnnotation](http://gvs.gs.washington.edu/SeattleSeqAnnotation)). All variants within the coding regions of the 9 risk loci (*ABCA7* [NM\_019112]; *CD2AP* [NM\_012120]; *MS4A6A* [NM\_152851]; *CR1* [NM\_000573]; *BIN1* [NM\_139343]; *PICALM* [NM\_001206946]; *EPHA1* [NM\_005232]; *CLU* [NM\_001831]; *CD33* [NM\_001772] have been collected and analysed. Indels were excluded from the merged dataset because not targeted in the ADNI subcohort (**Figure 1**). (Further details are provided in the supplementary materials).

## 2.5 Statistical Analysis

In the single-variant analysis, allele frequencies were calculated for each low frequency and rare coding variants in cases and controls and Fisher's exact test on allelic association was performed. To study the joint effect of the variants detected, we performed a gene-based analysis with SKAT and c-alpha test and we analysed together for each gene the whole spectrum of allelic variability (common, low frequency, rare, coding and non-coding).

C-alpha test and SKAT are closely related, being both non-burden test, analyzing and collapsing the effect of genetic variants of different frequency (common and rare), effect (protective, damaging and neutral) and effect size (modest, moderate, strong). SKAT can be considered an expansion of the c-alpha test because overcomes some of its limits. Indeed, SKAT 1) can be applied also to the study of continuous traits and 2) do not need any permutation.

Low frequency and rare variants were defined as having a  $1\% < \text{MAF} < 5\%$  and  $\text{MAF} < 1\%$ , respectively, either in cases or controls.

All computations, C-alpha and SKAT tests were performed in R (version x64 3.0.2, <http://www.r-project.org/>) and PLINK/SEQ.

A p-value of 0.05 was set as a nominal significance threshold. Based on multiple testing correction, the thresholds for single variant and gene-based association tests are defined by  $p\text{-value} = 1.25 \times 10^{-3}$  ( $0.05/40$  [total number of coding low frequency and rare variants detected in our study]) and  $5.5 \times 10^{-3}$  ( $0.05/9\text{genes}$ ), respectively. Furthermore, we excluded singletons from the single-variant analysis, since a variant observed only once is not largely informative about the overall distribution (Neale et al., 2011). However, we pooled the singletons together and analysed their collective effect in the gene-based analysis (SKAT and c-alpha test).

In addition, we report the complete list of coding variants detected in these GWAS loci in the supplementary table (**Table S1**).

### 3. RESULTS

The discovery set consisted of a total of 332 sporadic and mainly late-onset AD cases and 676 elderly controls of British and North American ancestry (**Table 1**).

A total of 289 single nucleotide variants (SNVs) were identified. Among these, 128 (44.3%) were nonsynonymous, 72 (24.9%) were synonymous, 83 (28.7%) UTR and 6 (2%) were intronic variants. Among the missense variants, 99 (77.34%) were rare ( $MAF < 1\%$ ), and 72.72% of these (72) were singletons (a variant observed only once either in cases or controls). 15 nonsynonymous variants (11.7%) were low frequency ( $1\% < MAF < 5\%$ ) and 16 (12.5%) were common ( $MAF \geq 5\%$ ). In addition, we report 14 novel coding variant (not reported in ExAC, released 13 January 2015, or dbSNP 137). None of the detected low frequency and rare coding variants cluster within common haplotypes ( $MAF \geq 5\%$ ) and therefore could have been missed by GWASs and chip based fine-mapping approaches. (**Table S1, Table S2**). We report the presence of two or three low frequency and/or rare variants in the studied genes in the same individual, both for cases and controls (**Table S3a and Table S3b**).

Overall, the total variant frequency of the 9 GWAS loci in our study was in line with the one reported for the European-American cohort in the Exome Variant Server (EVS) (<http://evs.gs.washington.edu/EVS/>). The only exception was represented by *CR1*, that showed a 2.7-fold higher relative frequency of total variants, compared to the EVS database. (**Table S4**).

*PICALM* harbors the lowest burden of low frequency and rare coding variants (3.27 coding variants per kb of coding sequence). By contrast, *CD33*, presents the highest relative frequency of coding variants and the lowest relative frequency of damaging variants (9.14 and 0.91 coding and damaging variants, respectively), suggesting that the majority of coding variability in *CD33* is likely non-functional (**Table S5**).

*BIN* and *ABCA7* display the highest relative proportion of damaging variants (3.92 [87.3% of its coding variability] and 3.72 [60% of its coding variability] damaging coding variants per kb of coding sequence, respectively), thus arguing for a potential functional impact of

missense mutations at these loci (**Table S5**). Moreover, *ABCA7* was the only gene harboring nonsense mutations.

Most of the low frequency and rare coding variability identified within these loci exerts generally a relatively modest effect (mean OR=1.1) that is comparable to those observed for common and generally noncoding variants identified by GWASs (**Table S1**).

### 3.1 Single coding variant association test

The main hits of the single-variants association test map mainly to *ABCA7* (**Table 2**). Particularly, we report *ABCA7* p.G215S (rs72973581), that was the only low-frequency (MAF=4.3%) missense variant showing a trend towards significance in the single-marker association test (p-value =0.02 and corrected p-value=0.8) in the discovery set and was statistically significant after Bonferroni correction (p-value =  $6 \times 10^{-4}$  and corrected p-value=0.024) in the combined datasets (discovery set and follow-up dataset).

Rs72973581 [A] results in a glycine to serine amino acid change at the position 215 of *ABCA7* (G215S) and its frequency was 1.56-fold higher in controls compared to cases (MAF= 4.66 and 7.24 for cases and controls, respectively), arguing for a protective effect (OR=0.6, 95% CI=0.38-0.95). This variant was present in homozygosity in one control. The study possessed relatively low power to detect a significant association between cases and controls for low frequency and rare variants. Therefore, we have followed up *ABCA7* p.G215S, carrying out Sanger sequencing in an independent dataset composed of 307 Caucasian late-onset AD cases and 501 elderly Caucasian controls (p-value= 0.012; OR= 0.54, 95% CI 0.31-0.89). In this follow-up dataset, we confirmed a higher frequency of the *ABCA7* p.G215S variant in controls compared to cases (carrier frequency= 13.5% vs 7.8% [1.7-fold], MAF= 7% vs 4.3% [1.6-fold], respectively).

Finally, we report also a common coding polymorphism in *ABCA7* (p.R1349Q, rs3745842) in our cohort, that map 1.3 kb from a reported GWAS hit, rs3752246, but cluster within a

different common haplotype block (MAF>5%)(Naj et al., 2011)(**Table S2**). Rs3745842 major allele (G) was more frequent in cases compared to controls, although the association was nominally significant after multiple testing correction (p-value=  $1.4 \times 10^{-3}$ , corrected p-value=0.081) (**Table S6**).

### 3.2 Loss of function mutations in *ABCA7*

Loss of function mutations in *ABCA7* have been recently reported as main mechanism explaining the GWAS signal and the increased susceptibility to AD.

In our cohort, we detected 5 loss of function mutations in *ABCA7*: 2 stopgain mutations (p.Y1579X and p.E1974X) and 3 splice-site or near splice-site mutations (c.-7-2 A>G, c.-7-7 T>C and c.231-12C>A). *ABCA7* p.E1974X and c.231-12C>A are novel variants and, together with p.Y1579X, are singletons, detected only in controls (**Table S7**).

Importantly, the enrichment for *ABCA7* LoF mutations and novel variants in controls did not rely on the sequencing strategy (exome sequencing vs genome sequencing). These variants have not been indeed mainly detected in the 199 BYU controls, that underwent genome sequencing (**Table S8**)

Moreover, 3 very rare indels have been identified in controls in the NIH-UCL cohort (p.1402delT, p.1638delCTT and p.1749delCTACTG). *ABCA7* p.1749delCTACTG is a novel mutation and *ABCA7* p.1402delT was present also in one case. These indels have been excluded from the pooled dataset because not targeted in the ADNI subcohort (**Table S9**) (**Figure 1**).

Finally, 3 nonsense mutations (p.W749X, p.W903X and p.R1754X) and one splice-site mutation, (c.4416+2T>G) did not pass either the sample or variant quality control (QC) criteria and therefore have not been included in the study (**Table S10**).

### 3.3 Gene-based association test

In addition to single-marker analysis, we carried out gene-wide analysis to combine the joint signal from multiple variants (coding variants and flanking UTRs) within a gene and to

provide greater statistical power than that for single-marker tests. All the variants (nonsynonymous, synonymous, UTRs, singletons) located within the studied genes and their exon-intron flanking regions were collapsed together and their combined effect was studied. *ABCA7* was the main hit both in the SKAT and c-alpha test, nominally and statistically significant, respectively, after multiple testing correction (corrected p-value= 0.6 and  $5.3 \times 10^{-3}$  respectively) (**Tables 3a and 3b**). Importantly, given the exclusion of indels in the merged discovery dataset, the presence of only 2 nonsense singleton mutations in *ABCA7* (p.Y1579X and p.E1974X, both detected in controls) and 4 putative splice-site mutations (rs3752229, rs2242437, c.231-12C>A and rs182233998, the latter one nominally significant in controls), our findings are not influenced by a burden of LoF mutations in *ABCA7*. Considering the very rare frequency of these LoF variants, their detection with sufficient power would have required a very large sample size. By contrast, the top signals are represented mainly by common-low frequency coding variants with an higher frequency in controls compared to cases and with a modest to intermediate protective effect ( $0.329 < OR < 0.755$ ) (**Table S6**).

#### 4. DISCUSSION

We report the results of single-variant and gene-based association tests performed in *BIN1*, *CLU*, *CR1*, *PICALM*, *MS4A6A*, *ABCA7*, *EPHA1*, *CD33*, *CD2AP* in a cohort composed of 332 apparently sporadic and mainly late-onset AD cases and 676 elderly Caucasian controls from North America and UK. In the single-variant association test we have analysed the effect of low frequency and rare coding variants (MAF<5%), aiming to identify potential functional variant(s) underlying the GWAS hit(s). In the gene-based analysis (SKAT and c-alpha test), we collapsed the full spectrum of variants identified in these loci to study their collective effect.

We do not report any pathogenic mutation in *APP*, *PSEN1* and *PSEN2* in our cohort. However, one of the controls was an heterozygous carrier of the protective variant *APP* p.A673T (MAF  $7 \times 10^{-4}$  in our cohort and MAF  $5 \times 10^{-4}$  among the European non-finnish, ExAC database, released 13 January 2015)(Jonsson et al., 2012)

*TREM2* p.R47H, the second most common risk factor for sporadic AD, has been detected in 6 cases (1.8%) and 4 controls (0.59%) and, likely given our small sample size, with a MAF=0.2%, was not significantly associated to AD (p-value= 0.09).

*ABCA7* was the only significant hit in the c-alpha test and harbors a low-frequency coding variant (p.G215S, rs72973581), whose minor allele confers a modest (OR=0.57, 95% CI= 0.41-0.80) but statistically significant protection (corrected p-value =0.024) against AD. Importantly, this SNP was not present in several GWAS or exome SNP arrays and does not cluster within common haplotypes identified by tagging SNPs, whereas it has been detected through *ABCA7* direct sequencing in the current study. Therefore, rs72973581 would have stayed likely undetected using common fine-mapping genotyping arrays. In addition, it does not cluster in the risk haplotypes identified by GWAS main hits (rs3764650, rs115550680, rs3752246, rs4147929), suggesting an independent signal and a likely different pathogenic mechanism of the major allele (Hollingworth et al., 2011)(Reitz et al., 2013)(Naj et al., 2011)(Liu et al., 2014). Importantly, the *ABCA7* p.G215S significant protective role against AD is supported by a targeted resequencing study of *ABCA7* in a Belgian cohort, where rs72973581 (A) frequency was 1.34-fold higher in controls compared to cases (p-value= 0.055) (Cuyvers et al., 2015). Notably, the main variant associated to LOAD in this Belgian cohort was a low frequency intronic variant (rs78117248) that did not pass our QC filter. However, in line with our findings, Cuyvers et al. report an enrichment for common and low frequency polymorphism with a modest protective role in *ABCA7*. Importantly, among the top 10 genetic variants identified in our study, 3 missense mutations (rs74176364, rs114782266, rs117187003) have been described associated also to autism spectrum disorder (ASD), strongly pointing towards a functional role of these amino acid changes and suggesting a possible shared pathogenic mechanisms underpinning neurodegenerative and neurodevelopmental diseases (He et al., 2014).

Interestingly, several lines of evidence reported that a significant decrease in *ABCA7* levels is associated to AD. At this regard, different and likely not mutually exclusive mechanisms have been described to influence the protein level: 1) common and generally non-coding variants in regulatory regions; 2) alternative splicing; 3) increased CpG island methylation

(Humphries et al., 2015)(Vasquez et al., 2013). Recently, also LoF mutations in *ABCA7* have been shown to significantly increase the susceptibility to AD in the Islandic population (Steinberg et al., 2015). This has been replicated in two different populations (Caucasian North American and Belgian) by two independent studies (Vardarajan et al., 2015)(Cuyvers et al., 2015). Therefore we report another potential mechanism, through which low frequency protein coding variability in *ABCA7* may influence AD risk.

Notably, *ABCA7* p.G215S provides critical insights into the genetic architecture of diseases, reinforcing the view that GWAS loci, likewise Mendelian genes, harbor low frequency and rare protective coding variants that can counteract with a similar effect size the damaging alleles (OR  $\approx 0.6$  vs  $\approx 1.1$  and  $\approx 0.2$  vs  $\approx 5$ , for GWAS loci and Mendelian genes, respectively) (Nejentsev et al., 2009)(Rivas et al., 2011)(Jonsson et al., 2012)(Asante et al., 2015).

*ABCA7* is mainly expressed in leukocytes and in myelo-lymphatic tissues (thymus, spleen and bone marrow) and microglia in the brain ([http://web.stanford.edu/group/barres\\_lab/brain\\_rnaseq.html](http://web.stanford.edu/group/barres_lab/brain_rnaseq.html))(<http://www.uniprot.org/>) (Kim et al., 2008). *ABCA7* encodes for ATP-binding cassette sub-family A member 7 (*ABCA7*), a multi-pass protein, present on the cell, Golgi and endosome membranes (<http://www.uniprot.org/>). *In vitro* and *in vivo* experiments have shown *ABCA7* pivotal role in phagocytosis and a likely modest role in HDL biogenesis. In *Abca7*<sup>-/-</sup> mice, macrophages and microglia display impaired phagocytosis and clearance of amyloid from the brain, which leads to cognitive impairment (Iwamoto et al., 2006)(Tanaka et al., 2011).

Therefore, *ABCA7*, likewise *TREM2* and *CD33*, may play an important role in regulating microglial uptake and clearance of A $\beta$  debris.

*ABCA7* p.G215S clusters within the extracellular topological domain of *ABCA7*. Remarkably, at the homologous residue, the serine is the reference amino acid in *ABCA7* in different mammals and in the homologous protein *ABCA1* in humans. Thus, suggesting that this amino acid may confer some biologic advantage and may have been positively selected during the evolution (**Figure S1 and S2**).



Likewise other low frequency and rare protective variants at the GWAS loci (**Table S11**), *ABCA7* p.G215 is a relatively conserved residue among different species (**Figure S1**) and this amino acid change (glycine to serine) may only slightly modify the protein activity (-5.86, 56, Gerd and Grantham score, respectively). Moreover, it has been reported as a tolerated change and benign, arguing against any possible loss of function or significant impairment of *ABCA7*, that have been indeed associated to increased risk for AD (Steinberg et al., 2015). The biological effect of this substitution may therefore lead to a mild *ABCA7* gain of function, possibly strengthening the interaction with a binding protein or regulating its expression. Although *ABCA7* p.G215 has not been predicted to be a coding target for miRNA (<https://www.umm.uni-heidelberg.de>), a possible post-transcriptional or post-translational regulation should not be excluded. Importantly, the substitution of a glycine with a serine may imply an additional substrate for serine-kinases or proteases. Moreover, in a similar way, *ABCA1*, whose LoF variants have been associated to AD (Nordestgaard et al., 2015)(Kim et al., 2012), has been reported to be particularly enriched for low frequency and rare coding variants with an average 1.5-fold higher frequency in controls compared to LOAD cases and a modest protective effect in a Greek cohort (OR =0.96-0.38) (Lupton et al., 2014) (**Table S12**).

Thus, understanding the effect of *ABCA7* p.G215S has the potential of unravelling new pathogenic mechanisms underpinning AD and may provide a promising therapeutic target that would not significantly alter *ABCA7* overall physiological function, which is critical for AD development.

Finally, we support Vardarajan et al. resequencing study of the GWAS loci (Vardarajan et al., 2015), confirming a burden of damaging variants in *ABCA7* and *BIN1* (**Table S5**) and to a lesser extent in *CD2AP*, *EPHA1* and *MS4A6A* (main hits in the gene-based analysis) (**Table 3a and 3b**), highlighting their potential role as susceptibility loci for LOAD.

However, we could not replicate the main hits detected by Vardarajan et al. in the single-variant analysis, either because such variants have been targeted but not detected in our cohort (*ABCA7* p.E1679X, *EPHA1* p.P460L and *BIN1* p.K358R) or because the variants have been targeted but eliminated by the QC filter (*CD2AP* p.K633R). Thus, suggesting a possible

lack of replication compared to the previous studies attributable to the different population but also different sequencing strategies, capture and coverage. Nevertheless, *EPHA1* and *CD2AP* harbor 2 of the main hits detected in the single-variant analysis in our cohort (rs11768549 and rs143297472, respectively)(**Table 2**), with rs11768549 already associated with the rapid progression of the disease in a cohort of Caucasian North American LOAD cases (Wang et al., 2015).

In summary, we support previous studies, suggesting that 1) *ABCA7* significantly influences AD risk; 2) *ABCA7* p.G215S is likely to reduce the susceptibility to AD; 3) GWAS hits are pleomorphic loci harboring a complex spectrum of variants synergistically contributing to the disease phenotype with different mechanisms, effects (damaging, protective and neutral) and effect sizes ( $0 < OR < 4$ ) and 4) gene-based approaches are effective methods to mine genetic data and to accurately filter potential candidate genes.

## ACKNOWLEDGEMENTS

We thank contributors, including the Alzheimer's Disease Centers who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible.

This study was supported by the Alzheimer's Research UK, the Medical Research Council (MRC), the Wellcome Trust/MRC Joint Call in Neurodegeneration Award (WT089698) to the UK Parkinson's Disease Consortium (whose members are from the University College London Institute of Neurology, the University of Sheffield, and the MRC Protein Phosphorylation Unit at the University of Dundee), grants (P50 AG016574, U01 AG006786, and R01 AG18023), the National Institute for Health Research Biomedical Research Unit in Dementia at University College London Hospitals, University College London; an anonymous donor, the Big Lottery (to Dr. Morgan); a fellowship from Alzheimer's Research UK (to Dr. Guerreiro); and the Intramural Research Programs of the National Institute on Aging and the National Institute of Neurological Disease and Stroke, National Institutes of Health (Department of Health and Human Services Project number, ZO1 AG000950-10). The MRC

London Neurodegenerative Diseases Brain Bank and the Manchester Brain Bank from Brains for Dementia Research are jointly funded from ARUK and AS. This work was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Department of Health and Human Services, project number ZO1 AG000950-10. Samples from the National Cell Repository for Alzheimer's Disease (NCRAD), which receives government support under a cooperative agreement grant (U24 AG21886) awarded by the National Institute on Aging (NIA), were used in this study. NIH grant R01 AG042611 to Kauwe J.

## CONFLICT OF INTEREST STATEMENT STATEMENT

The authors declare no competing financial or personal interests that can influence the presented work

## REFERENCES

- Asante, E.A., Smidak, M., Grimshaw, A., Houghton, R., Tomlinson, A., Jeelani, A., Jakubcova, T., Hamdan, S., Richard-Londt, A., Linehan, J.M., Brandner, S., Alpers, M., Whitfield, J., Mead, S., Wadsworth, J.D.F., Collinge, J., 2015. A naturally occurring variant of the human prion protein completely prevents prion disease. *Nature* 522, 478–481. doi:10.1038/nature14510
- Beaudoin, M., Goyette, P., Boucher, G., Lo, K.S., Rivas, M.A., Stevens, C., Alikashani, A., Ladouceur, M., Ellinghaus, D., Törkvist, L., Goel, G., Lagacé, C., Annese, V., Bitton, A., Begun, J., Brant, S.R., Bresso, F., Cho, J.H., Duerr, R.H., Halfvarson, J., McGovern, D.P.B., Radford-Smith, G., Schreiber, S., Schumm, P.L., Sharma, Y., Silverberg, M.S., Weersma, R.K., Quebec IBD Genetics Consortium, NIDDK IBD Genetics Consortium, International IBD Genetics Consortium, D'Amato, M., Vermeire, S., Franke, A., Lettre, G., Xavier, R.J., Daly, M.J., Rioux, J.D., 2013. Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. *PLoS Genet.* 9, e1003723. doi:10.1371/journal.pgen.1003723
- Chartier-Harlin, M.C., Parfitt, M., Legrain, S., Pérez-Tur, J., Brousseau, T., Evans, A., Berr, C., Vidal, O., Roques, P., Gourlet, V., 1994. Apolipoprotein E, epsilon 4 allele as a major risk factor for sporadic early and late-onset forms of Alzheimer's disease: analysis of the 19q13.2 chromosomal region. *Hum. Mol. Genet.* 3, 569–574.
- Cuyvers, E., De Roeck, A., Van den Bossche, T., Van Cauwenberghe, C., Bettens, K., Vermeulen, S., Mattheijssens, M., Peeters, K., Engelborghs, S., Vandenbulcke, M., Vandenberghe, R., De Deyn, P.P., Van Broeckhoven, C., Sleegers, K., 2015. Mutations

- in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. *Lancet Neurol.* 14, 814–822. doi:10.1016/S1474-4422(15)00133-7
- Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga, C., Sassi, C., Kauwe, J.S.K., Younkin, S., Hazrati, L., Collinge, J., Pocock, J., Lashley, T., Williams, J., Lambert, J.-C., Amouyel, P., Goate, A., Rademakers, R., Morgan, K., Powell, J., St George-Hyslop, P., Singleton, A., Hardy, J., Alzheimer Genetic Analysis Group, 2013. TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.* 368, 117–127. doi:10.1056/NEJMoa1211851
- Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Williams, A., Jones, N., Thomas, C., Stretton, A., Morgan, A.R., Lovestone, S., Powell, J., Proitsi, P., Lupton, M.K., Brayne, C., Rubinsztein, D.C., Gill, M., Lawlor, B., Lynch, A., Morgan, K., Brown, K.S., Passmore, P.A., Craig, D., McGuinness, B., Todd, S., Holmes, C., Mann, D., Smith, A.D., Love, S., Kehoe, P.G., Hardy, J., Mead, S., Fox, N., Rossor, M., Collinge, J., Maier, W., Jessen, F., Schürmann, B., Heun, R., van den Bussche, H., Heuser, I., Kornhuber, J., Wiltfang, J., Dichgans, M., Frölich, L., Hampel, H., Hüll, M., Rujescu, D., Goate, A.M., Kauwe, J.S.K., Cruchaga, C., Nowotny, P., Morris, J.C., Mayo, K., Sleegers, K., Bettens, K., Engelborghs, S., De Deyn, P.P., Van Broeckhoven, C., Livingston, G., Bass, N.J., Gurling, H., McQuillin, A., Gwilliam, R., Deloukas, P., Al-Chalabi, A., Shaw, C.E., Tsolaki, M., Singleton, A.B., Guerreiro, R., Mühleisen, T.W., Nöthen, M.M., Moebus, S., Jöckel, K.-H., Klopp, N., Wichmann, H.-E., Carrasquillo, M.M., Pankratz, V.S., Younkin, S.G., Holmans, P.A., O'Donovan, M., Owen, M.J., Williams, J., 2009. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.* 41, 1088–1093. doi:10.1038/ng.440
- He, Z., O'Roak, B.J., Smith, J.D., Wang, G., Hooker, S., Santos-Cortez, R.L.P., Li, B., Kan, M., Krumm, N., Nickerson, D.A., Shendure, J., Eichler, E.E., Leal, S.M., 2014. Rare-variant extensions of the transmission disequilibrium test: application to autism exome sequence data. *Am. J. Hum. Genet.* 94, 33–46. doi:10.1016/j.ajhg.2013.11.021
- Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J.-C., Carrasquillo, M.M., Abraham, R., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Jones, N., Stretton, A., Thomas, C., Richards, A., Ivanov, D., Widdowson, C., Chapman, J., Lovestone, S., Powell, J., Proitsi, P., Lupton, M.K., Brayne, C., Rubinsztein, D.C., Gill, M., Lawlor, B., Lynch, A., Brown, K.S., Passmore, P.A., Craig, D., McGuinness, B., Todd, S., Holmes, C., Mann, D., Smith, A.D., Beaumont, H., Warden, D., Wilcock, G., Love, S., Kehoe, P.G., Hooper, N.M., Vardy, E.R.L.C., Hardy, J., Mead, S., Fox, N.C., Rossor, M., Collinge, J., Maier, W., Jessen, F., Rütger, E., Schürmann, B., Heun, R., Kölsch, H., van den Bussche, H., Heuser, I., Kornhuber, J., Wiltfang, J., Dichgans, M., Frölich, L., Hampel, H., Gallacher, J., Hüll, M., Rujescu, D., Giegling, I., Goate, A.M., Kauwe, J.S.K., Cruchaga, C., Nowotny, P., Morris, J.C., Mayo, K., Sleegers, K., Bettens, K., Engelborghs, S., De Deyn, P.P., Van Broeckhoven, C., Livingston, G., Bass, N.J., Gurling, H., McQuillin, A., Gwilliam, R., Deloukas, P., Al-Chalabi, A., Shaw, C.E., Tsolaki, M., Singleton, A.B., Guerreiro, R., Mühleisen, T.W., Nöthen, M.M., Moebus, S., Jöckel, K.-H., Klopp, N., Wichmann, H.-E., Pankratz, V.S., Sando, S.B., Aasly, J.O., Barcikowska, M., Wszolek, Z.K., Dickson, D.W., Graff-Radford, N.R., Petersen, R.C.,

- Alzheimer's Disease Neuroimaging Initiative, van Duijn, C.M., Breteler, M.M.B., Ikram, M.A., DeStefano, A.L., Fitzpatrick, A.L., Lopez, O., Launer, L.J., Seshadri, S., CHARGE consortium, Berr, C., Campion, D., Epelbaum, J., Dartigues, J.-F., Tzourio, C., Alperovitch, A., Lathrop, M., EADI1 consortium, Feulner, T.M., Friedrich, P., Riehle, C., Krawczak, M., Schreiber, S., Mayhaus, M., Nicolhaus, S., Wagenpfeil, S., Steinberg, S., Stefansson, H., Stefansson, K., Snaedal, J., Björnsson, S., Jonsson, P.V., Chouraki, V., Genier-Boley, B., Hiltunen, M., Soininen, H., Combarros, O., Zelenika, D., Delepine, M., Bullido, M.J., Pasquier, F., Mateo, I., Frank-Garcia, A., Porcellini, E., Hanon, O., Coto, E., Alvarez, V., Bosco, P., Siciliano, G., Mancuso, M., Panza, F., Solfrizzi, V., Nacmias, B., Sorbi, S., Bossù, P., Piccardi, P., Arosio, B., Annoni, G., Seripa, D., Pilotto, A., Scarpini, E., Galimberti, D., Brice, A., Hannequin, D., Licastro, F., Jones, L., Holmans, P.A., Jonsson, T., Riemenschneider, M., Morgan, K., Younkin, S.G., Owen, M.J., O'Donovan, M., Amouyel, P., Williams, J., 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat. Genet.* 43, 429–435. doi:10.1038/ng.803
- Humphries, C., Kohli, M.A., Whitehead, P., Mash, D.C., Pericak-Vance, M.A., Gilbert, J., 2015. Alzheimer disease (AD) specific transcription, DNA methylation and splicing in twenty AD associated loci. *Mol. Cell. Neurosci.* 67, 37–45. doi:10.1016/j.mcn.2015.05.003
- Iwamoto, N., Abe-Dohmae, S., Sato, R., Yokoyama, S., 2006. ABCA7 expression is regulated by cellular cholesterol through the SREBP2 pathway and associated with phagocytosis. *J. Lipid Res.* 47, 1915–1927. doi:10.1194/jlr.M600127-JLR200
- Jonsson, T., Atwal, J.K., Steinberg, S., Snaedal, J., Jonsson, P.V., Björnsson, S., Stefansson, H., Sulem, P., Gudbjartsson, D., Maloney, J., Hoyte, K., Gustafson, A., Liu, Y., Lu, Y., Bhangale, T., Graham, R.R., Huttenlocher, J., Björnsdóttir, G., Andreassen, O.A., Jönsson, E.G., Palotie, A., Behrens, T.W., Magnusson, O.T., Kong, A., Thorsteinsdóttir, U., Watts, R.J., Stefansson, K., 2012. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 488, 96–99. doi:10.1038/nature11283
- Kim, J., Yoon, H., Ramírez, C.M., Lee, S.-M., Hoe, H.-S., Fernández-Hernando, C., Kim, J., 2012. MiR-106b impairs cholesterol efflux and increases A $\beta$  levels by repressing ABCA1 expression. *Exp. Neurol.* 235, 476–483. doi:10.1016/j.expneurol.2011.11.010
- Kim, W.S., Weickert, C.S., Garner, B., 2008. Role of ATP-binding cassette transporters in brain lipid transport and neurological disease. *J. Neurochem.* 104, 1145–1166. doi:10.1111/j.1471-4159.2007.05099.x
- Lambert, J.-C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., Combarros, O., Zelenika, D., Bullido, M.J., Tavernier, B., Letenneur, L., Bettens, K., Berr, C., Pasquier, F., Fiévet, N., Barberger-Gateau, P., Engelborghs, S., De Deyn, P., Mateo, I., Franck, A., Helisalmi, S., Porcellini, E., Hanon, O., European Alzheimer's Disease Initiative Investigators, de Pancorbo, M.M., Lendon, C., Dufouil, C., Jaillard, C., Leveillard, T., Alvarez, V., Bosco, P., Mancuso, M., Panza, F., Nacmias, B., Bossù, P., Piccardi, P., Annoni, G., Seripa, D., Galimberti, D., Hannequin, D., Licastro, F., Soininen, H., Ritchie, K., Blanché, H., Dartigues, J.-F., Tzourio, C., Gut, I., Van Broeckhoven, C., Alperovitch, A., Lathrop, M., Amouyel, P., 2009. Genome-wide association study

- identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat. Genet.* 41, 1094–1099. doi:10.1038/ng.439
- Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., Russo, G., Thorton-Wells, T.A., Jones, N., Smith, A.V., Chouraki, V., Thomas, C., Ikram, M.A., Zelenika, D., Vardarajan, B.N., Kamatani, Y., Lin, C.F., Gerrish, A., Schmidt, H., Kunkle, B., Dunstan, M.L., Ruiz, A., Bihoreau, M.T., Choi, S.H., Reitz, C., Pasquier, F., Cruchaga, C., Craig, D., Amin, N., Berr, C., Lopez, O.L., De Jager, P.L., Deramecourt, V., Johnston, J.A., Evans, D., Lovestone, S., Letenneur, L., Morón, F.J., Rubinsztein, D.C., Eiriksdottir, G., Sleegers, K., Goate, A.M., Fiévet, N., Huentelman, M.W., Gill, M., Brown, K., Kamboh, M.I., Keller, L., Barberger-Gateau, P., McGuinness, B., Larson, E.B., Green, R., Myers, A.J., Dufouil, C., Todd, S., Wallon, D., Love, S., Rogaeva, E., Gallacher, J., St George-Hyslop, P., Clarimon, J., Lleo, A., Bayer, A., Tsuang, D.W., Yu, L., Tsolaki, M., Bossù, P., Spalletta, G., Proitsi, P., Collinge, J., Sorbi, S., Sanchez-Garcia, F., Fox, N.C., Hardy, J., Deniz Naranjo, M.C., Bosco, P., Clarke, R., Brayne, C., Galimberti, D., Mancuso, M., Matthews, F., European Alzheimer's Disease Initiative (EADI), Genetic and Environmental Risk in Alzheimer's Disease, Alzheimer's Disease Genetic Consortium, Cohorts for Heart and Aging Research in Genomic Epidemiology, Moebus, S., Mecocci, P., Del Zompo, M., Maier, W., Hampel, H., Pilotto, A., Bullido, M., Panza, F., Caffarra, P., Nacmias, B., Gilbert, J.R., Mayhaus, M., Lannefelt, L., Hakonarson, H., Pichler, S., Carrasquillo, M.M., Ingelsson, M., Beekly, D., Alvarez, V., Zou, F., Valladares, O., Younkin, S.G., Coto, E., Hamilton-Nelson, K.L., Gu, W., Razquin, C., Pastor, P., Mateo, I., Owen, M.J., Faber, K.M., Jonsson, P.V., Combarros, O., O'Donovan, M.C., Cantwell, L.B., Soininen, H., Blacker, D., Mead, S., Mosley, T.H., Bennett, D.A., Harris, T.B., Fratiglioni, L., Holmes, C., de Bruijn, R.F., Passmore, P., Montine, T.J., Bettens, K., Rotter, J.I., Brice, A., Morgan, K., Foroud, T.M., Kukull, W.A., Hannequin, D., Powell, J.F., Nalls, M.A., Ritchie, K., Lunetta, K.L., Kauwe, J.S., Boerwinkle, E., Riemschneider, M., Boada, M., Hiltunen, M., Martin, E.R., Schmidt, R., Rujescu, D., Wang, L.S., Dartigues, J.F., Mayeux, R., Tzourio, C., Hofman, A., Nöthen, M.M., Graff, C., Psaty, B.M., Jones, L., Haines, J.L., Holmans, P.A., Lathrop, M., Pericak-Vance, M.A., Launer, L.J., Farrer, L.A., van Duijn, C.M., Van Broeckhoven, C., Moskvina, V., Seshadri, S., Williams, J., Schellenberg, G.D., Amouyel, P., 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* 45, 1452–1458. doi:10.1038/ng.2802
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* 25, 1754–1760. doi:10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- Liu, G., Li, F., Zhang, S., Jiang, Y., Ma, G., Shang, H., Liu, J., Feng, R., Zhang, L., Liao, M., Zhao, B., Li, K., 2014. Analyzing large-scale samples confirms the association between the ABCA7 rs3764650 polymorphism and Alzheimer's disease susceptibility. *Mol. Neurobiol.* 50, 757–764. doi:10.1007/s12035-014-8670-4



- Lohmueller, K.E., Sparsø, T., Li, Q., Andersson, E., Korneliussen, T., Albrechtsen, A., Banasik, K., Grarup, N., Hallgrimsdottir, I., Kiil, K., Kilpeläinen, T.O., Krarup, N.T., Pers, T.H., Sanchez, G., Hu, Y., Degiorgio, M., Jørgensen, T., Sandbæk, A., Lauritzen, T., Brunak, S., Kristiansen, K., Li, Y., Hansen, T., Wang, J., Nielsen, R., Pedersen, O., 2013. Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. *Am. J. Hum. Genet.* 93, 1072–1086. doi:10.1016/j.ajhg.2013.11.005
- Lupton, M.K., Proitsi, P., Lin, K., Hamilton, G., Daniilidou, M., Tsolaki, M., Powell, J.F., 2014. The role of ABCA1 gene sequence variants on risk of Alzheimer's disease. *J. Alzheimers Dis. JAD* 38, 897–906. doi:10.3233/JAD-131121
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110
- Naj, A.C., Jun, G., Beecham, G.W., Wang, L.-S., Vardarajan, B.N., Buross, J., Gallins, P.J., Buxbaum, J.D., Jarvik, G.P., Crane, P.K., Larson, E.B., Bird, T.D., Boeve, B.F., Graff-Radford, N.R., De Jager, P.L., Evans, D., Schneider, J.A., Carrasquillo, M.M., Ertekin-Taner, N., Younkin, S.G., Cruchaga, C., Kauwe, J.S.K., Nowotny, P., Kramer, P., Hardy, J., Huentelman, M.J., Myers, A.J., Barmada, M.M., Demirci, F.Y., Baldwin, C.T., Green, R.C., Rogaeva, E., St George-Hyslop, P., Arnold, S.E., Barber, R., Beach, T., Bigio, E.H., Bowen, J.D., Boxer, A., Burke, J.R., Cairns, N.J., Carlson, C.S., Carney, R.M., Carroll, S.L., Chui, H.C., Clark, D.G., Corneveaux, J., Cotman, C.W., Cummings, J.L., DeCarli, C., DeKosky, S.T., Diaz-Arrastia, R., Dick, M., Dickson, D.W., Ellis, W.G., Faber, K.M., Fallon, K.B., Farlow, M.R., Ferris, S., Frosch, M.P., Galasko, D.R., Ganguli, M., Gearing, M., Geschwind, D.H., Ghetti, B., Gilbert, J.R., Gilman, S., Giordani, B., Glass, J.D., Growdon, J.H., Hamilton, R.L., Harrell, L.E., Head, E., Honig, L.S., Hulette, C.M., Hyman, B.T., Jicha, G.A., Jin, L.-W., Johnson, N., Karlawish, J., Karydas, A., Kaye, J.A., Kim, R., Koo, E.H., Kowall, N.W., Lah, J.J., Levey, A.I., Lieberman, A.P., Lopez, O.L., Mack, W.J., Marson, D.C., Martiniuk, F., Mash, D.C., Masliah, E., McCormick, W.C., McCurry, S.M., McDavid, A.N., McKee, A.C., Mesulam, M., Miller, B.L., Miller, C.A., Miller, J.W., Parisi, J.E., Perl, D.P., Peskind, E., Petersen, R.C., Poon, W.W., Quinn, J.F., Rajbhandary, R.A., Raskind, M., Reisberg, B., Ringman, J.M., Roberson, E.D., Rosenberg, R.N., Sano, M., Schneider, L.S., Seeley, W., Shelanski, M.L., Slifer, M.A., Smith, C.D., Sonnen, J.A., Spina, S., Stern, R.A., Tanzi, R.E., Trojanowski, J.Q., Troncoso, J.C., Van Deerlin, V.M., Vinters, H.V., Vonsattel, J.P., Weintraub, S., Welsh-Bohmer, K.A., Williamson, J., Woltjer, R.L., Cantwell, L.B., Dombroski, B.A., Beekly, D., Lunetta, K.L., Martin, E.R., Kamboh, M.I., Saykin, A.J., Reiman, E.M., Bennett, D.A., Morris, J.C., Montine, T.J., Goate, A.M., Blacker, D., Tsuang, D.W., Hakonarson, H., Kukull, W.A., Foroud, T.M., Haines, J.L., Mayeux, R., Pericak-Vance, M.A., Farrer, L.A., Schellenberg, G.D., 2011. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* 43, 436–441. doi:10.1038/ng.801
- Neale, B.M., Rivas, M.A., Voight, B.F., Altshuler, D., Devlin, B., Orho-Melander, M., Kathiresan, S., Purcell, S.M., Roeder, K., Daly, M.J., 2011. Testing for an unusual distribution of rare variants. *PLoS Genet.* 7, e1001322. doi:10.1371/journal.pgen.1001322

- Nejentsev, S., Walker, N., Riches, D., Egholm, M., Todd, J.A., 2009. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 324, 387–389. doi:10.1126/science.1167728
- Nordestgaard, L.T., Tybjærg-Hansen, A., Nordestgaard, B.G., Frikke-Schmidt, R., 2015. Loss-of-function mutation in ABCA1 and risk of Alzheimer's disease and cerebrovascular disease. *Alzheimers Dement. J. Alzheimers Assoc.* doi:10.1016/j.jalz.2015.04.006
- Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., De, T., UK Brain Expression Consortium, North American Brain Expression Consortium, Coin, L., de Silva, R., Cookson, M.R., Singleton, A.B., Hardy, J., Ryten, M., Weale, M.E., 2014. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* 17, 1418–1428. doi:10.1038/nn.3801
- Reitz, C., Jun, G., Naj, A., Rajbhandary, R., Vardarajan, B.N., Wang, L.-S., Valladares, O., Lin, C.-F., Larson, E.B., Graff-Radford, N.R., Evans, D., De Jager, P.L., Crane, P.K., Buxbaum, J.D., Murrell, J.R., Raj, T., Ertekin-Taner, N., Logue, M., Baldwin, C.T., Green, R.C., Barnes, L.L., Cantwell, L.B., Fallin, M.D., Go, R.C.P., Griffith, P., Obisesan, T.O., Manly, J.J., Lunetta, K.L., Kamboh, M.I., Lopez, O.L., Bennett, D.A., Hendrie, H., Hall, K.S., Goate, A.M., Byrd, G.S., Kukull, W.A., Foroud, T.M., Haines, J.L., Farrer, L.A., Pericak-Vance, M.A., Schellenberg, G.D., Mayeux, R., Alzheimer Disease Genetics Consortium, 2013. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E  $\epsilon$ 4, and the risk of late-onset Alzheimer disease in African Americans. *JAMA* 309, 1483–1492. doi:10.1001/jama.2013.2973
- Rivas, M.A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C.K., Boucher, G., Ripke, S., Ellinghaus, D., Burt, N., Fennell, T., Kirby, A., Latiano, A., Goyette, P., Green, T., Halfvarson, J., Haritunians, T., Korn, J.M., Kuruvilla, F., Lagacé, C., Neale, B., Lo, K.S., Schumm, P., Törkvist, L., National Institute of Diabetes and Digestive Kidney Diseases Inflammatory Bowel Disease Genetics Consortium (NIDDK IBDGC), United Kingdom Inflammatory Bowel Disease Genetics Consortium, International Inflammatory Bowel Disease Genetics Consortium, Dubinsky, M.C., Brant, S.R., Silverberg, M.S., Duerr, R.H., Altshuler, D., Gabriel, S., Lettre, G., Franke, A., D'Amato, M., McGovern, D.P.B., Cho, J.H., Rioux, J.D., Xavier, R.J., Daly, M.J., 2011. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet.* 43, 1066–1073. doi:10.1038/ng.952
- Service, S.K., Teslovich, T.M., Fuchsberger, C., Ramensky, V., Yajnik, P., Koboldt, D.C., Larson, D.E., Zhang, Q., Lin, L., Welch, R., Ding, L., McLellan, M.D., O'Laughlin, M., Fronick, C., Fulton, L.L., Magrini, V., Swift, A., Elliott, P., Jarvelin, M.-R., Kaakinen, M., McCarthy, M.I., Peltonen, L., Pouta, A., Bonnycastle, L.L., Collins, F.S., Narisu, N., Stringham, H.M., Tuomilehto, J., Ripatti, S., Fulton, R.S., Sabatti, C., Wilson, R.K., Boehnke, M., Freimer, N.B., 2014. Re-sequencing expands our understanding of the phenotypic impact of variants at GWAS loci. *PLoS Genet.* 10, e1004147. doi:10.1371/journal.pgen.1004147
- Seshadri, S., Fitzpatrick, A.L., Ikram, M.A., DeStefano, A.L., Gudnason, V., Boada, M., Bis, J.C., Smith, A.V., Carassquillo, M.M., Lambert, J.C., Harold, D., Schrijvers, E.M.C., Ramirez-Lorca, R., Debette, S., Longstreth, W.T., Janssens, A.C.J.W., Pankratz, V.S., Dartigues, J.F., Hollingworth, P., Aspelund, T., Hernandez, I., Beiser, A., Kuller, L.H., Koudstaal, P.J., Dickson, D.W., Tzourio, C., Abraham, R., Antunez, C., Du, Y., Rotter, J.I.,



- Aulchenko, Y.S., Harris, T.B., Petersen, R.C., Berr, C., Owen, M.J., Lopez-Arrieta, J., Varadarajan, B.N., Becker, J.T., Rivadeneira, F., Nalls, M.A., Graff-Radford, N.R., Campion, D., Auerbach, S., Rice, K., Hofman, A., Jonsson, P.V., Schmidt, H., Lathrop, M., Mosley, T.H., Au, R., Psaty, B.M., Uitterlinden, A.G., Farrer, L.A., Lumley, T., Ruiz, A., Williams, J., Amouyel, P., Younkin, S.G., Wolf, P.A., Launer, L.J., Lopez, O.L., van Duijn, C.M., Breteler, M.M.B., CHARGE Consortium, GERAD1 Consortium, EADI1 Consortium, 2010. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 303, 1832–1840. doi:10.1001/jama.2010.574
- Steinberg, S., Stefansson, H., Jonsson, T., Johannsdottir, H., Ingason, A., Helgason, H., Sulem, P., Magnusson, O.T., Gudjonsson, S.A., Unnsteinsdottir, U., Kong, A., Helisalmi, S., Soininen, H., Lah, J.J., DemGene, Aarsland, D., Fladby, T., Ulstein, I.D., Djurovic, S., Sando, S.B., White, L.R., Knudsen, G.-P., Westlye, L.T., Selbæk, G., Giegling, I., Hampel, H., Hiltunen, M., Levey, A.I., Andreassen, O.A., Rujescu, D., Jonsson, P.V., Bjornsson, S., Snaedal, J., Stefansson, K., 2015. Loss-of-function variants in *ABCA7* confer risk of Alzheimer's disease. *Nat. Genet.* doi:10.1038/ng.3246
- Tanaka, N., Abe-Dohmae, S., Iwamoto, N., Fitzgerald, M.L., Yokoyama, S., 2011. HMG-CoA reductase inhibitors enhance phagocytosis by upregulating ATP-binding cassette transporter A7. *Atherosclerosis* 217, 407–414. doi:10.1016/j.atherosclerosis.2011.06.031
- Vardarajan, B.N., Ghani, M., Kahn, A., Sheikh, S., Sato, C., Barral, S., Lee, J.H., Cheng, R., Reitz, C., Lantigua, R., Reyes-Dumeyer, D., Medrano, M., Jimenez-Velazquez, I.Z., Rogaeva, E., St George-Hyslop, P., Mayeux, R., 2015. Rare coding mutations identified by sequencing of Alzheimer disease genome-wide association studies loci. *Ann. Neurol.* 78, 487–498. doi:10.1002/ana.24466
- Vasquez, J.B., Fardo, D.W., Estus, S., 2013. *ABCA7* expression is associated with Alzheimer's disease polymorphism and disease status. *Neurosci. Lett.* 556, 58–62. doi:10.1016/j.neulet.2013.09.058
- Visscher, P.M., Brown, M.A., McCarthy, M.I., Yang, J., 2012. Five years of GWAS discovery. *Am. J. Hum. Genet.* 90, 7–24. doi:10.1016/j.ajhg.2011.11.029
- Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164. doi:10.1093/nar/gkq603
- Wang, X., Lopez, O.L., Sweet, R.A., Becker, J.T., DeKosky, S.T., Barmada, M.M., Demirci, F.Y., Kamboh, M.I., 2015. Genetic determinants of disease progression in Alzheimer's disease. *J. Alzheimers Dis. JAD* 43, 649–655. doi:10.3233/JAD-140729

**Figure 1. Pipeline of our study design.** WES, whole exome sequencing; WGS, whole genome sequencing; INDELS, in frame insertions and deletions; SNVs, single nucleotide variants; CTRLs, controls; UTRs, untranslated regions. \* INDELS have been excluded from the discovery cohort because not targeted in the ADNI dataset.

| COHORTS                                      | N   | TYPE      | SEQUENCING STRATEGY | ORIGIN  | AGE (YRS)<br>MEAN $\pm$ SD(RANGE) | MALE<br>(%) | APOE<br>E4+ (%) |
|--|-----|-----------|---------------------|---|-----------------------------------|-------------|-----------------|
| DISCOVERY SET                                |     |           |                     |   |                                   |             |                 |
| <b>NIH-UCL</b>                               |     |           |                     |   |                                   |             |                 |
| cases  | 127 | neuropath | Exome sequencing    | Caucasian (British)   | 65.5(41-94)                       | 46.4        | 58              |
| controls                                     | 204 | neuropath | Exome sequencing    | Caucasian (British, North American)   | 79.8 (61-102)                     | 58.3        | 45              |
| <b>WashU</b>                                 |     |           |                     |   |                                   |             |                 |
| cases  | 23  | clinical  | Exome sequencing    | Caucasian (North American)  | 57 (46-75)                        | 52.17       | NA              |
| controls                                     | 16  | clinical  | Exome sequencing    | Caucasian (North American)  | 79.5 (75-92)                      | 43.7        | NA              |
| <b>ADNI</b>                                  |     |           |                     |   |                                   |             |                 |
| cases  | 182 | clinical  | Exome sequencing    | Caucasian (North American)  | 74.65 (55-90)                     | 67          | 56.6            |
| controls                                     | 257 | clinical  | Exome sequencing    | Caucasian (North American)  | 74.68 (60-90)                     | 50.1        | 27.6            |
| <b>BYU</b>                                   |     |           |                     |   |                                   |             |                 |
| controls                                     | 199 | clinical  | Genome sequencing   | Caucasian (North American)  | 80.8 (75-94.5)                    | 37.7        | 100             |
| FOLLOW-UP GENOTYPING SET<br>ABCA7 rs72973581 |     |           |                     |   |                                   |             |                 |
| <b>NIH-NIA</b>                               |     |           |                     |   |                                   |             |                 |
| cases  | 307 | clinical  | Sanger sequencing   | Caucasian (North American, British, Dutch, Italian, Portuguese)                   | average >65y                      |             |                 |
| controls                                     | 501 | clinical  | Sanger sequencing   | Caucasian, (North American, British, Greek, German, Polish, Australian, Canadian) | >60y                              |             |                 |

**Table 1. Description of the different cohorts used in this study.** N, number; WES, whole exome sequencing; WGS, whole genome sequencing; YRS, years.

| GENE  | POSITION    | MA | cDNA change | Aa change            | rs          | MAF cases-ctrls (%)    | MAF ExAC (%) | SIFT        | POLYPHEN          | Mutation assessor | aa/Aa/AA cases        | aa/Aa/AA controls       | p-value         | Corr. p-value | OR (95% CI)                                 |
|-------|-------------|----|-------------|----------------------|-------------|------------------------|--------------|-------------|-------------------|-------------------|-----------------------|-------------------------|-----------------|---------------|---|
| ABCA7 | 19:1043103  | A  | c.G643A     | p.G215S              | rs72973581  | 4.66-7.24 <sup>§</sup> | 4.31***      | Tolerated   | Benign            | low               | 0/31/301<br>0/55/584* | 1/96/579<br>1/164/1012* | 0.02<br>0.0006* | 0.8<br>0.024  | 0.61<br>(0.38-0.95)<br>0.57*<br>(0.41-0.80) |
| ABCA7 | 19:1050996  | A  | c.G2629A    | p.A877T**            | rs74176364  | 0.3-1.18               | 1.69         | Deleterious | Benign            | low               | 0/2/330               | 0/16/660                | 0.07            | 2.8           | 0.25<br>(0.02-1.07)                         |
| EPHA1 | 7:143095153 | A  | c.G1475A    | p.R492Q <sup>#</sup> | rs11768549  | 2.56-1.47              | 1.21         | tolerated   | benign            |                   | 0/17/315              | 1/18/657                | 0.07            | 2.8           | 1.86<br>(0.89-3.84)                         |
| ABCA7 | 19:1059056  | A  | c.G5435A    | p.R1812H**           | rs114782266 | 1.5-0.81               | 1.05         | Tolerated   | Benign            | neutral           | 0/10/322              | 0/11/665                | 0.16            | 6.4           | 1.87<br>(0.70-4.92)                         |
| ABCA7 | 19:1057343  | A  | c.G4795A    | p.V1599M**           | rs117187003 | 0.6-0.22               | 0.3          | Deleterious | Possibly damaging | medium            | 0/4/328               | 0/3/673                 | 0.22            | 8.8           | 2.73<br>(0.45-18.7)                         |
| CD2AP | 6:47573971  | A  | c.G1488A    | p.M496I              | rs143297472 | 0.3-0.07               | NA           | Tolerated   | Benign            |                   | 0/2/330               | 0/1/675                 | 0.25            | 10            | 4.08<br>(0.21-241.3)                        |
| ABCA7 | 19:1047537  | C  | c.A2153C    | p.N718T              | rs3752239   | 1.65-2.44              | 7.02         | Deleterious | Benign            | low               | 0/11/321              | 0/33/641                | 0.32            | 12.8          | 0.66<br>(0.29-1.37)                         |

**Table2. Most significant variants detected in our discovery set.** Position is in hg19/GRCh37. MAF, minor allele frequency; OR, odds ratio; CI=confidence interval. Corr, Corrected p-value, p-value after Bonferroni correction (p-value\*40 [number of variants considered in the single-variant association test])

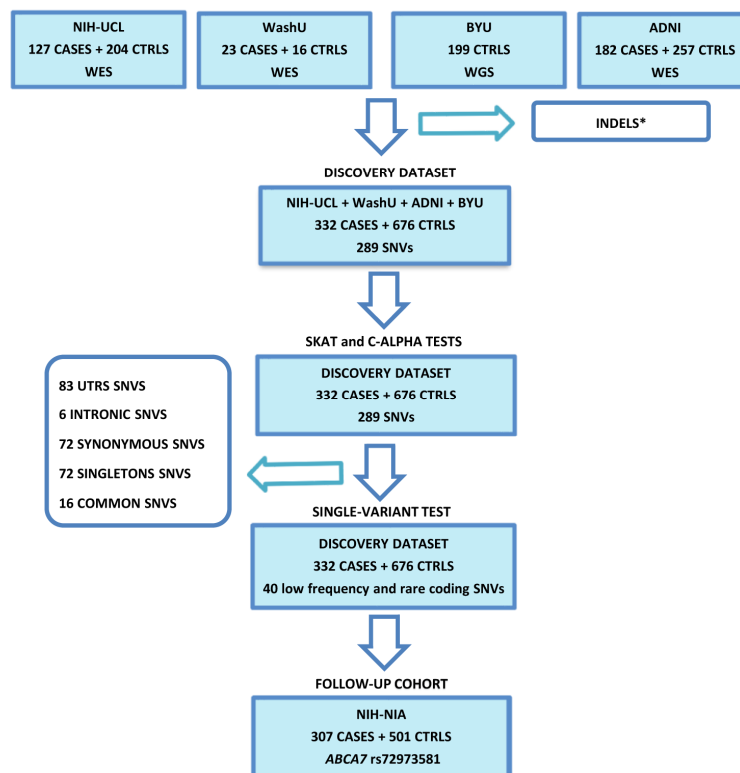
\*Combined results discovery and follow-up data set. <sup>§</sup>MAF cases-ctrls reported a Belgian cohort = 4.66-6.27(Cuyvers et al., 2015) \*\*Variants reported associated also with autism spectrum disorders (ASD) (He et al., 2014). \*\*\*MAF in ExAC (European non Finnish)= 6.14% and MAF in EVS (European American)= 6.24%. <sup>#</sup> Variant reported associated to a more rapid disease progression

| TRANSCRIPT ID | POSITION                  | GENE          | N.VARIANTS | TEST   | P-VALUE     | CORRECTED P-VALUE |
|---------------|---------------------------|---------------|------------|--------|-------------|-------------------|
| NM_019112     | chr19:1040131..1065563    | <i>ABCA7</i>  | 72         | CALPHA | 0.000590992 | 0.0053            |
| NM_012120     | chr6:47445789..47594915   | <i>CD2AP</i>  | 20         | CALPHA | 0.0353261   | 0.31              |
| NM_152851     | chr11:59939123..59950523  | <i>MS4A6A</i> | 11         | CALPHA | 0.0548523   | 0.49              |
| NM_000573     | chr1:207669709..207814864 | <i>CR1</i>    | 72         | CALPHA | 0.0677083   | 0.6               |
| NM_139343     | chr2:127805799..127864546 | <i>BIN1</i>   | 27         | CALPHA | 0.0730337   | 0.65              |
| NM_001206946  | chr11:85668697..85779900  | <i>PICALM</i> | 19         | CALPHA | 0.0742857   | 0.66              |
| NM_005232     | chr7:143088365..143105830 | <i>EPHA1</i>  | 30         | CALPHA | 0.106557    | 0.95              |
| NM_001831     | chr8:27454493..27472251   | <i>CLU</i>    | 29         | CALPHA | 0.444444    | 3.99              |
| NM_001772     | chr19:51728380..51743144  | <i>CD33</i>   | 13         | CALPHA | 0.714286    | 6.42              |

**Table 3a. Results from the c-alpha test performed.** Position is in hg19/GRCh37. Corrected p-value, p-value after Bonferroni correction ( $p\text{-value} \times 9[\text{number of genes considered in the single-gene based analysis}]$ ).

| TRANSCRIPT ID | POSITION                  | GENE          | N.VARIANTS | TEST | P-VALUE   | CORRECTED P-VALUE |
|---------------|---------------------------|---------------|------------|------|-----------|-------------------|
| NM_019112     | chr19:1040131..1065563    | <i>ABCA7</i>  | 72         | SKAT | 0.0737249 | 0.66              |
| NM_005232     | chr7:143088365..143105830 | <i>EPHA1</i>  | 30         | SKAT | 0.298155  | 2.68              |
| NM_139343     | chr2:127805799..127864546 | <i>BIN1</i>   | 27         | SKAT | 0.447218  | 4.02              |
| NM_012120     | chr6:47445789..47594915   | <i>CD2AP</i>  | 20         | SKAT | 0.448924  | 4.04              |
| NM_000573     | chr1:207669709..207814864 | <i>CR1</i>    | 72         | SKAT | 0.510539  | 4.59              |
| NM_001831     | chr8:27454493..27472251   | <i>CLU</i>    | 29         | SKAT | 0.59029   | 5.31              |
| NM_152851     | chr11:59939123..59950523  | <i>MS4A6A</i> | 11         | SKAT | 0.937765  | 8.43              |
| NM_001772     | chr19:51728380..51743144  | <i>CD33</i>   | 13         | SKAT | 0.93899   | 8.45              |
| NM_001206946  | chr11:85668697..85779900  | <i>PICALM</i> | 19         | SKAT | 0.943777  | 8.49              |

**Table 3b. Results from the SKAT test performed.** Position is in hg19/GRCh37. Corrected p-value, p-value after Bonferroni correction (p-value\*9[number of genes considered in the single-gene based analysis]).



# *ABCA7 p.G215S as potential protective factor for Alzheimer's disease*

Sassi C.<sup>1, 2, 3, 4</sup>, Nalls M.A.<sup>2</sup>, Ridge P.G.<sup>5</sup>, Gibbs R.<sup>2</sup>, Ding J.<sup>2</sup>, Lupton M.K.<sup>6</sup>, Troakes C.<sup>6</sup>, Lunnon K.<sup>6</sup>, Al-Sarraj S.<sup>6</sup>, Brown K.S.<sup>7</sup>, Medway C.<sup>7</sup>, Clement N.<sup>7</sup>, Lord J.<sup>7</sup>, Turton J.<sup>7</sup>, Bras J.<sup>1</sup>, Almeida MR.<sup>8</sup>, ARUK Consortium\*, Holstege H.<sup>9</sup>, Louwersheimer E.<sup>9</sup>, van der Flier W.<sup>9</sup>, Scheltens P.<sup>9</sup>, Van Swieten J.C.<sup>10</sup>, Santana I.<sup>11</sup>, Oliveira C.<sup>12</sup>, Morgan K.<sup>5</sup>, Powell J.F.<sup>4</sup>, Kauwe J.S.<sup>3</sup>, Cruchaga C.<sup>13</sup>, Goate AM.<sup>14</sup>, Singleton A.B.<sup>2</sup>, Guerreiro R.<sup>1</sup>, Hardy J.<sup>1</sup>

\*The Alzheimer's Research UK (ARUK) Consortium: Peter Passmore, David Craig, Janet Johnston, Bernadette McGuinness, Stephen Todd, Queen's University Belfast, UK; Reinhard Heun, Royal Derby Hospital, UK; Heike Kölsch, University of Bonn, Germany; Patrick G. Kehoe, University of Bristol, UK; Emma R.L.C. Vardy, Salford Royal NHS Foundation Trust, UK; Nigel M. Hooper, David M. Mann, Stuart Pickering-Brown, University of Manchester, UK; Kristelle Brown, James Lowe, Kevin Morgan, University of Nottingham, UK; A. David Smith, Gordon Wilcock, Donald Warden, University of Oxford (OPTIMA), UK; Clive Holmes, University of Southampton, UK.

## Highlights

- Genome-wide association studies (GWASs) have been effective approaches to investigate common genetic variability underlying complex traits in a systematic and unbiased way.
- Despite the evidence showing the contribution of these loci to AD pathogenesis, their genetic architecture has not been extensively investigated.
- We have used exome and genome sequencing data to analyse the single independent and joint effect of rare and low frequency protein coding variants in 9 AD GWAS loci with the strongest effect sizes after *APOE* (*BIN1*, *CLU*, *CR1*, *PICALM*, *MS4A6A*, *ABCA7*, *EPHA1*, *CD33*, *CD2AP*) in a cohort of 332 sporadic AD cases and 676 elderly controls of British and North American ancestry.
- We identified coding variability in *ABCA7* as contributing to AD risk. This locus harbors a low frequency coding variant (p.G215S, rs72973581, MAF=4.3%) conferring a modest but statistically significant protection against AD (p-value=  $6 \times 10^{-4}$ , OR= 0.57, 95% CI 0.41-0.80).
- Notably, our results are not driven by an enrichment of loss of function variants in *ABCA7*, recently reported as main pathogenic factor underlying AD risk at this locus.
- Our study confirms the role of *ABCA7* in AD and provide new insights that should address functional studies.